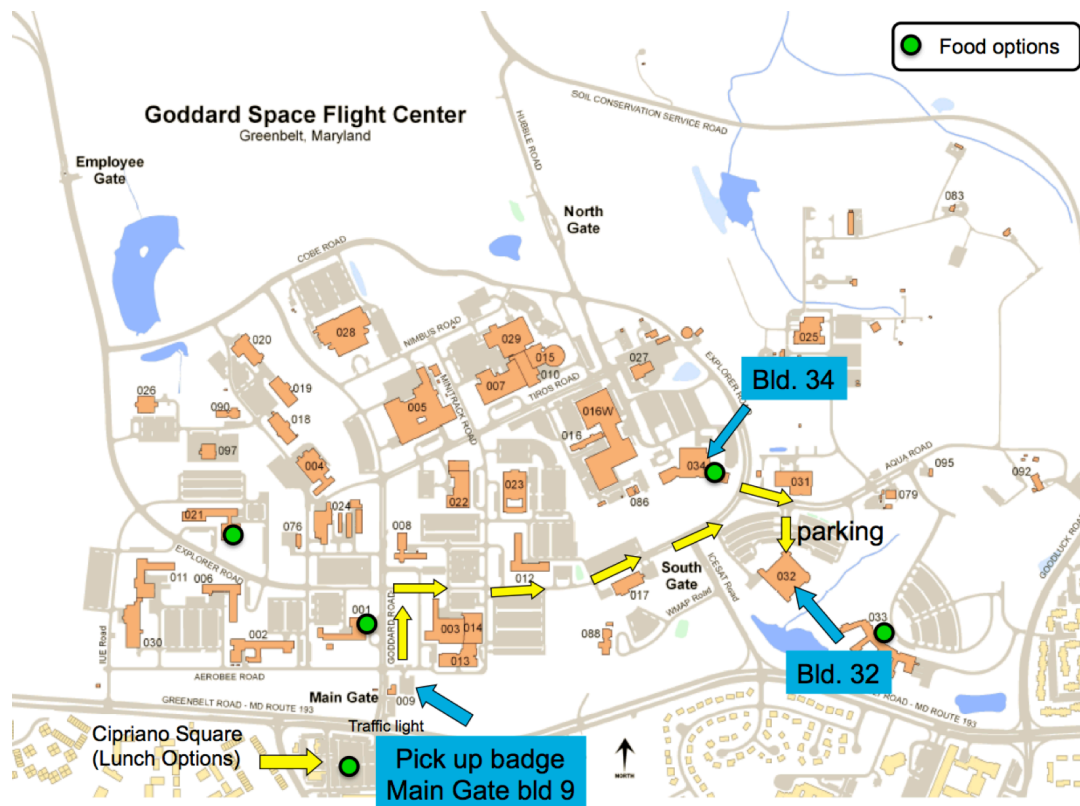


Metadata Curation Summit 2 - Agenda and Meeting Minutes

Where is this meeting? We will be in building 32 (B32), room N202.



Agenda

May 15, 2017

Time	Topic	Presenter
9:00 - 9:45 AM	Introductions/Icebreaker	Dana
9:45 AM - 10:00 AM	Goals	Kaylin and Katie
10:00 AM - 10:15 AM	Overview of ARC Process and #curation	Kaylin and Jeanne'
10:15 AM - 10:45 AM	Preview of Dashboard	E84
10:45 AM - 11:05 AM	Deep Dive of ARC Reports and Rule	Kaylin and Adam
11:05 AM - 11:15 AM	ARC Schedule	Adam
11:15 AM - 11:25 AM	BREAK	
11:25 AM - 11:45 AM	Keywords: Curation vs. Creation Perspectives	Jeanne'
11:45 AM - 12:05 PM	Keywords: Change Process and Version 8.5 Release	Steve

12:05 PM - 1:00 PM	LUNCH	
1:00 PM - 1:15 PM	CMR Tools	Dana
1:15 PM - 1:30 PM	Related URLs and Online Resource	Jeanne' and Kaylin
1:30 PM - 2:30 PM	Group Breakout	All
2:30 PM - 2:45 PM	BREAK	
2:45 PM - 2:55 PM	Platform/Sensor/Instrument	Group A
2:55 PM - 3:05 PM	Spatial Extent	Group A
3:05 PM - 3:15 PM	Spatial Information: Datum, etc.	Group A
3:15 PM - 3:25 PM	Version ID	Group B
3:25 PM - 3:35 PM	Citation	Group B
3:35 PM - 3:45 PM	Processing Level	Group B
3:45 PM - 3:55 PM	Collection Progress Enums	Erich
3:55 PM - 4:15 PM	Enumerations	Kathy
4:15 PM - 4:30 PM	DOI	Dana
As Time Permits	Additional Metadata Topics (License, Related URL, Online Resource, Org / Data Contacts, Long Name, Org Names)	Varied
4:45 PM - 5:00 PM	Action Item Review	Dana

Link to presentations: <https://drive.google.com/drive/folders/0B202wR4N3yalR1dJcWNwbEZxOVU?usp=sharing>

Attendees:

Kaylin Bugbee (UAH/MSFC)

Jeanne le Roux (UAH/MSFC)

Adam Sisco (UAH/MSFC)

Steve Wharton (GCMD)

Tyler Stevens (GCMD)

Asad Ullah (MODAPS/LAADS)

Bhaskar Ramachandran (MODAPS/VIIRS)

Katie Baynes (ESDIS)

Bruce Vollmer (GES DISC)

Dana Ostenga (GES DISC)

Carolyn Gacke (LP DAAC)
Simon Cantrell (CMR)
Kathy Carr (CMR)
Erich Reiter (CMR)
Carlee Loeser (GES DISC)
James Johnson (GES DISC)
Jon Pals (SDPS/ECS)
Chris Finch (PODAAC)
Chris Lynnes (ESDIS)
Rahul Ramachandran (GHRC)
Valerie Dixon (CMR)
Andrew Baker (Element84)
Scott Caltagirone (Element84)
Neil Divine (MODAPS)
Jeanne Laurencelle (ASF)
Alicia Scott (OBDAAC)

Notes:

Overview Presentation/Schedule/Processes

- Question: Does #curation channel have user feedback?
- Answer: It has feedback from user services representatives, but that is all
- Question: What do you do if there is no clear consensus on a #curation polls?
- Answer: Katie Baynes said "Kaylin - have an opinion and make the call!"
- Question: Does curation workoff impact DAAC workplans?
- Answer: If the DAAC foresees an impact to current work plan, they need to raise that to ESDIS

- ☐ [Kaylin Bugbee](#) - Request was made to create an easy to read document on the metadata curation rules (meanwhile on the #curation channel - a request was made to set up a process to facilitate rule changes or new rules; also for a centralized document location)
- ☐ [Kathleen Carr](#) - Share the mapping of Earthdata search client values to UMM-JSON fields as well as indicating whether humanizers are used with this community once it is complete.

- Also it was mentioned that we should educate on how decisions are made for the metadata rules and reviews
- Suggestion was made to offer QA tips when entering metadata into MMT (similar to what GCMD currently offers)

Dashboard Presentation (Scott and Andrew)

- ☒ Dana Shum - Discuss "static map" and potential use of tags with Kaylin and dashboard team
- ☐ [Kaylin Bugbee](#) Check on script output - 'Flagged by Script' what gets checked in that box? Just red issues? Need to follow up on this.
- Question: Business rules - how do you keep one reviewer from undoing another review while in progress?
- Answer: Part of future work
- Also has rules that you can't mark something as complete if a second review is needed or sections have not been reviewed
- For reporting - make sure the color category is easy to understand in the high level reporting (red, yellow, blue)
- 3 sections of the dashboard -

1. Ingest a record
2. Review the record
3. Generate reports

- Question: What is the expectation from granule reviews?
- Answer: DAAC should make programmatic updates for their granules and reingest as necessary

- Question: Per provider, can you see how many records have problems in each element?
- Answer: Not yet, but its possible in the future

- Question: Is the dashboard team updating pyCMR in github?
- Answer: No, but other people are updating the automated checks and it's not yet getting into github

Parameter Mapping Tool

- ☐ Jeanne' le Roux - what exactly is in the lookup table? provide this information
- Question: is this only limited to CF variables?
- Answer: yes, but perhaps it could be expanded in the future.
- Comment on netCDF-4 - mostly not CF compliant. How will the tool handle this? Maybe use a humanizer mapping?

GCMD

- Question: Who is selecting keyword reviewers?
- Answer: ESDIS Standards office facilitates keyword review - identifies qualified reviewers

- Question: Are 8.5 keywords available in docBuilder?
- Answer: Yes. And also in MMT. And they are also coming out in ARCs findings to the team.

- Recommendation for GCMD to fill gaps in keywords using standardized sources whenever possible. Example: AMS glossary
 - Answer: GCMD currently does this.
- Recommendation to investigate user search patterns (gave example of volcanism keywords) to discover good keywords
- Unbalanced facets - V8.5 was meant to resolve this (example given: atmospheric winds)
- Will move to yearly keyword releases - gave 2 weeks for reviewers, hope to give 4 weeks in the future since this will be a yearly activity

- Question: Shouldn't CMR update our keywords for us when they change?
- Answer: CMR's philosophy is to never change a provider's metadata. GCMD has a Change List Tool that can assist with DIF10 keyword changes. MMT is also currently building a Bulk Update functionality to assist with keyword changes.

- Bruce feels this is 'a little heavy handed' to ask DAACs to go in and make changes individually (understandably)

- ☒ Tyler Stevens Provide Kathleen Carr with information on the Platform/Instrument relationships in KMS
- ☒ Tyler Stevens Pursue offline the topics/actions not covered at the Curation Summit Meeting (Temporal Keywords)
- ☒ Tyler Stevens Correct workflow from the Keyword Forum to the keywords and back to the keywords
- ☒ Tyler Stevens Send Christopher Lynnes information about provider abbreviations in KMS

CMR Tools

- MMT has built in help text
- Bulk update workflow for keyword updates ready for beta testing in early July
- Need feedback from DAACs on how to make updates and changes easier related to UMM, etc...
- Native XML validation > UMM-C Validation > Business rule validation
- See slide on how to change CMR validations from 'warnings' to 'errors' for UMM-C compliance and keyword validation

Related URLs

- Collection Citation/Publication should be in 'Online Resource'
- Everything else should be in Related URLs

☐ Kaylin Bugbee Erich Reiter - What should relatedURLs look like in each dialect? Provide detailed examples

- Having one link which points as directly as possible to the described data is best but may not always be possible depending on the DAACs infrastructure
- Should a search tool be mapped as 'get data'? Is this the best way to describe this? Should be called a search tool not 'get data'

☒ Dana Shum - Write an EDSC ticket to not display RelatedURLs which would result in a circular use of Earthdata search (EDSC-1511)

- How is the search tool presenting this in front of the user - want to get users to get data and think about the description provided
- Question: Should browse be in the list of collection URLs (sample browse) or should you be allowed to only provide browse images at granule level?
- Answer: It's ok at the collection level, but not worth requiring
- Granule metadata should have data access link to directly download the granule as well as a link to the DOI/landing page
- OPeNDAP - IDV tool along with Panoply
- 'A link to the collection level OPeNDAP service' is just wrong unless you are aggregating stuff
- Provide a description for granules to help describe how to use OPeNDAP links? Need to provide more guidance on what these URLs do. Types and subtypes may provide clarity - strictly follow types and subtypes to specify
- Require the DAACs to configure OPeNDAP consistently (I think Chris L. is going to work on this?)
- Mime type
- These all need to be there with strong typing/subtyping and mime type
- ARC can develop recommendations on strong typing/subtyping

Version ID

- OB.DAAC - versioning by year (this is a nice idea). OB is doing new versions once a year and the version number updates to the current year number when re-processed
- LP.DAAC - ECS restriction of integer (not float)
- LAADS - uses float
- What is produced as 6.1 is put into an ECS DAAC as 61
- GESDISC - heavy weighting towards ECS method since there is a legacy of ECS
- What we need to understand - do all the number formats need to be the same?
- Version description used in all ECS records
- Also has implications for granules and may need to be changed
- The group likes this recommendation to standardize version but cautions that there are lots of moving pieces and could have major consequences for the DAACs
- What would be a good rollout strategy for a change of this magnitude? Archive system may have to change or accommodate this new information, UMM has to be modified, wait until bulk update tool is available
- Curation community develops recommendations and guidelines on converting from existing version ID to standardized version and opens up the recommendations for review
- Group recommends that DAACs make more extensive use of version description field
- Add another version element that migrates towards standardized version that can then be deprecated
- Start out with controlled vocabulary (algorithm version, data version, pge version, ?)

☐ Erich Reiter John Farley ProcessingVersion > Major/Minor should be integer

Resource Citation

- Everyone likes the idea of using crosscite.org
- We are ok with using the building blocks approach
- Dana Shum - Write ticket to have Earthdata search client display/link with crosscite.org
- Erich Reiter - Recommendation to split out names into first/middle/last and allow for multiple names. Determine if we have a ticket for this and if not, write it.
- There is the possibility for confusion/error providing information to DOI versus providing information in metadata
- About half of EOSDIS records have DOIs

Processing Level

- John Farley Need a definition for 1C or get rid of it.
- Humanizer to give a brief description of processing level
- Chris/John have done a humanizer mapping of short processing level descriptions (slide 37)
- LPDAAC and OBDAAC have taken an action to check on processing level

Instrument & Sensor

- Agree with items #1 and #2 from the slides
- Need a way to provide more information about Models and Computers. Not especially fond of 'computers'. Use of N/A is better.
- Consider having a place for describing the model run data
- Like the idea of eliminating sensor, strongly recommend validating relationship between platform and instrument
- Validate the first two levels (Platform/Instrument) but not 'composedOf'. Would need to give people lots of notice so that the metadata could be cleaned up first!

Spatial Extent/Bounding Rectangle

- Question: Can the spatial extent be generated by the CMR from the granule level records? (could also be temporal extent, but spatial is more helpful)
- Answer: In theory, yes. It's been discussed before but isn't on CMR's near term roadmap.
- The spatial extent/bounding rectangle should be more accurate

Spatial Information: Datum, etc....

- Please provide this information if you have it. Important to know and it isn't being submitted by providers much right now
- Not urgent to have a controlled vocabulary - more urgent to simply provide the information

☐ [Kaylin Bugbee](#) Provide more guidance or guidelines as needed for these terms

Collection Progress

- In Work -> Active
- Provisional -> Beta
- Deprecated -> Decommissioned
- Collection Maturity - This element is not ready for prime time
- There are 4 stages of validation with MODIS
- Note: Communication after the meeting points out that there are official NASA enums which we should use

DOI

- Can search by DOI now in CMR API
- ECHO10 mostly affected (DIF mapping already supports this)

☐

☒ [Erich Reiter](#) - Need to validate who "authority" is or if it is needed – UMM-C documentation in Jama is updated with guidance on the value of authority.

Enumerations

- ☐ [Erich Reiter](#) Post the enumeration somewhere it is easy to find (need work to figure out where and in what format)
- ☐ [Kaylin Bugbee](#) Rahul will consider what format is best for enums programmatically Long term - kms or some other service, JSON, CSV (what format)

Metadata forum/ FAQs

☐ [Kathleen Baynes](#) will send an invite out on how to join EOSDIS Slack